Name: Achyut Shukla Batch: ML - 1 PRN: 20070122005

॥वसुधैव कुटुम्बकम्॥

SYMBIOSIS INTERNATIONAL (DEEMED UNIVERSITY)

(Established under section 3 of the UGC Act 1956)

Re - accredited by NAAC with 'A' Grade

Founder: Prof.Dr. S. B. Mujumdar, M.Sc., Ph.D. (Awarded Padma Bhushan and Padma Shri by President of India)

Lab Assignment —6

Aim :

Implement simple Linear Regression and Logistic regression.

PART — A

Linear Regression Theory

:

1. Explain Linear models.

Linear models are a class of statistical and machine learning algorithms used for regression and classification tasks. They are based on the concept of a linear relationship between input features and the output. In these models, the goal is to find a linear equation that best describes the relationship between the input features and the target variable.

Types of Linear Models:

1. **Simple Linear Regression:** Involves a single input feature and aims to establish a linear relationship between that feature and the target variable.
2. **Multiple Linear Regression:** Handles multiple input features and predicts the target variable as a linear combination of all these features.
3. **Logistic Regression:** Despite its name, logistic regression is used for binary classification problems. It models the probability that an input instance belongs to a particular class using a logistic function, which ensures that the output is between 0 and 1.

2. Explain Linear Regression with solved example.

Linear regression finds the best-fitting line through the data points by minimizing the differences between predicted and actual values (residuals). The coefficients $w0$ and $w1$ are estimated to create this line, enabling predictions for new data points.

3. Discuss the different evaluation measure (MAE, RMSE etc.).

Mean Absolute Error (MAE): MAE calculates the average of the absolute differences between predicted and actual values. It measures the average magnitude of errors without considering their direction.
Root Mean Squared Error (RMSE): RMSE is similar to MAE but gives more weight to larger errors. It calculates the square root of the average of squared differences between predicted and actual values.
Mean Squared Error (MSE): MSE calculates the average of squared differences between predicted and actual values. It's closely related to RMSE and provides a way to measure the average squared error directly.
Coefficient of Determination (R-squared or $R2$): $R2$ measures the proportion of the variance in the dependent variable that's explained by the independent variables. It ranges from 0 to 1, where 0 indicates that the model doesn't explain any variability and 1 indicates a perfect fit.

Logistic regression

Theory :

1. Explain Logistic Regression with solved example.

Logistic Regression is a statistical and machine learning algorithm used for binary classification tasks. Despite its name, it's used to predict the probability of an instance belonging to a particular class rather than predicting numeric values as in linear regression. It's a widely used method for problems where the dependent variable is categorical and has two possible outcomes (usually represented as 0 and 1).

2. Discuss the different evaluation measure ((F-measures, Confusion Matrix etc.).).

Confusion Matrix: A confusion matrix is a table that describes the performance of a classification model. It breaks down the actual and predicted class labels into four categories: true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). From the confusion matrix, various metrics can be calculated.
F1-Score: The F1-score is the harmonic mean of precision and recall. It provides a balance between precision and recall and is particularly useful when the class distribution is imbalanced.

Receiver Operating Characteristic (ROC) Curve: The ROC curve is a graphical representation of the trade-off between sensitivity (recall) and specificity. It plots the true positive rate against the false positive rate for different threshold values.

PART — B

Experiment:

Implement linear and logistic regression and calculate the different evaluation measure.

References :

1. https://www.kaggle.com/datasets/sugandhkhobragade/bse30-daily-market-price-
20082018https://github.com/topics/naive-bayes-classifier
2. http://www.cs.toronto.edu/ delve/data/census-house/desc.html Linear

Regression Output:

1. Data Visualization

1. Training Set Visual

2. Test Set Visual

2. Evaluation Metrics

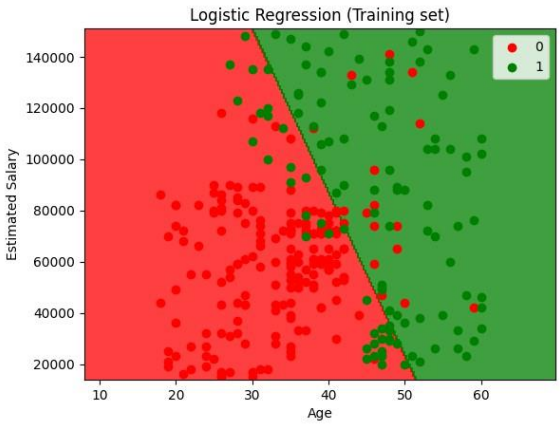
```
mse = mean_squared_error(y_test, y_pred)
rmse = np.sqrt(mse)
r2 = r2_score(y_test, y_pred)

# Display evaluation metrics
print("Mean Squared Error:", mse)
print("Root Mean Squared Error:", rmse)
print("R-squared:", r2)
```

```
Mean Squared Error: 21026037.329511296
Root Mean Squared Error: 4585.4157204675885
R-squared: 0.9749154407708353
```

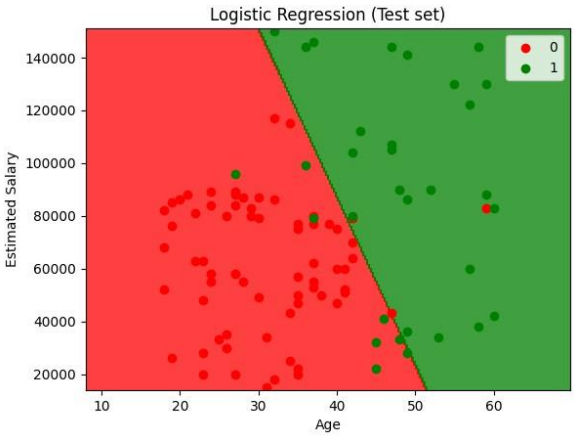
3. Evaluation Metric for Linear Regression

Logistic Regression Output:

1. Data Visualization

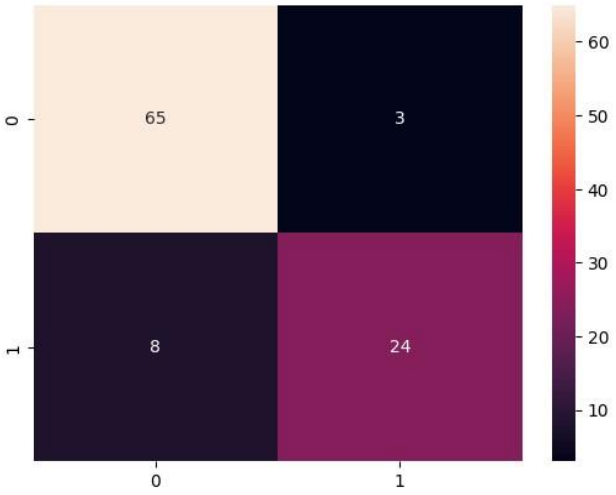


4. Training Set Visualization



5. Testing Set Visualization

2. Evaluation Metrics



6. Confusion Matrix

```
from sklearn.metrics import confusion_matrix, accuracy_score
cm = confusion_matrix(y_test, y_pred)
print(cm)
accuracy_score(y_test, y_pred)
```

```
[[65  3]
 [ 8 24]]
0.89
```

7. Accuracy

Inference Discussion

We have successfully implemented Linear Regression with the evaluation metrics of:- R Squared, RMSE & MSE. We have also implemented Logistic Regression with confusion matrix as our evaluation metric with accuracy.